

Justin A. Nelson (*Pro Hac Vice*)  
 Alejandra C. Salinas (*Pro Hac Vice*)  
 Collin Fredricks (*Pro Hac Vice*)  
**SUSMAN GODFREY L.L.P.**  
 1000 Louisiana Street, Suite 5100  
 Houston, TX 77002-5096  
 Telephone: (713) 651-9366  
 jnelson@susmangodfrey.com  
 asalinas@susmangodfrey.com  
 cfredricks@susmangodfrey.com

Rohit D. Nath (SBN 316062)  
**SUSMAN GODFREY L.L.P.**  
 1900 Avenue of the Stars, Suite 1400  
 Los Angeles, CA 90067-2906  
 Telephone: (310) 789-3100  
 RNath@susmangodfrey.com

Jordan W. Connors (*Pro Hac Vice*)  
**SUSMAN GODFREY L.L.P.**  
 401 Union Street, Suite 3000  
 Seattle, WA 98101  
 Telephone: (206) 516-3880  
 jconnors@susmangodfrey.com

J. Craig Smyser (*Pro Hac Vice*)  
**SUSMAN GODFREY L.L.P.**  
 One Manhattan West, 51st Floor,  
 New York, NY 10019  
 Telephone: (212) 336-8330  
 csmyser@susmangodfrey.com

*Proposed Co-Lead Counsel*

Rachel Geman (*Pro Hac Vice*)  
 Jacob S. Miller (*Pro Hac Vice*)  
 Danna Z. Elmasry (*Pro Hac Vice*)  
**LIEFF CABRASER HEIMANN  
 & BERNSTEIN, LLP**  
 250 Hudson Street, 8th Floor  
 New York, New York 10013-1413  
 Telephone: (212) 355-9500  
 rgeman@lchb.com  
 jmiller@lchb.com  
 delmasry@lchb.com

Daniel M. Hutchinson (SBN 239458)  
 Reilly T. Stoler (SBN 310761)  
**LIEFF CABRASER HEIMANN  
 & BERNSTEIN, LLP**  
 275 Battery Street, 29th Floor  
 San Francisco, CA 94111-3339  
 Telephone: (415) 956-1000  
 dhutchinson@lchb.com  
 rstoler@lchb.com

*Proposed Co-Lead Counsel*

Scott J. Sholder (*Pro Hac Vice*)  
 CeCe M. Cole (*Pro Hac Vice*)  
**COWAN DEBAETS ABRAHAMS  
 & SHEPPARD LLP**  
 60 Broad Street, 30th Floor  
 New York, New York 10010  
 Telephone: (212) 974-7474  
 ssholder@cdas.com  
 ccole@cdas.com

*Additional Counsel for the Class*

**UNITED STATES DISTRICT COURT  
 NORTHERN DISTRICT OF CALIFORNIA  
 SAN FRANCISCO DIVISION**

ANDREA BARTZ, ANDREA BARTZ,  
 INC., CHARLES GRAEBER, KIRK  
 WALLACE JOHNSON, and MJ + KJ, INC.,  
 individually and on behalf of others similarly  
 situated,

Plaintiffs,

v.

ANTHROPIC PBC,

Defendants

Case No. 3:24-cv-05417-WHA



**PLAINTIFFS' NOTICE OF MOTION AND  
 MOTION FOR CLASS CERTIFICATION**

**REDACTED**

## TABLE OF CONTENTS

		Page
1		
2		
3	I. INTRODUCTION .....	3
4	II. COMMON EVIDENCE OF HOW ANTHROPIC VALUES, COPIES, AND USES PLAINTIFFS' AND CLASS MEMBERS' COPYRIGHTED BOOKS.....	5
5	A. Anthropic Benefits from its LLMs, and Needs a High Volume of Books to Train and Develop Them. ....	5
6	B. Anthropic Deliberately Copied Books from Pirated Books Datasets. ....	6
7	C. Anthropic's [REDACTED] .....	8
8	D. Anthropic Copied Books to Create Training Datasets and Train and Develop its LLMs. ....	9
9	E. Anthropic's Use of Pirated and [REDACTED] Books Affected the Book- Purchase Market and the LLM-Licensing Market. ....	9
10	F. Plaintiffs Are Copyright Holders Who Bring a Common Copyright Act Claim. ....	10
11	III. LEGAL STANDARD .....	10
12	IV. THE PROPOSED CLASSES SATISFY RULE 23(a). ....	11
13	A. The Proposed Classes Are Objectively Defined. ....	11
14	B. The Proposed Classes Are Sufficiently Numerous. ....	11
15	C. Common Issues of Fact and Law Exist. ....	11
16	D. Plaintiffs' Claims Are Typical of the Classes' Claims. ....	12
17	E. Plaintiffs Fairly and Adequately Represent the Classes. ....	13
18	V. THE PROPOSED CLASSES SATISFY RULE 23(b)(2). ....	13
19	VI. THE PROPOSED CLASSES SATISFY RULE 23(b)(3). ....	14
20	A. Common Questions of Law and Fact Predominate. ....	14
21	1. Plaintiffs' Prima Facie Case of Copyright Infringement Will Rise or Fall on Common Proof. ....	15
22	2. Fair Use Will Be Resolved with Common Proof. ....	17
23	a. Pirated Books Class. ....	17
24	b. [REDACTED] Books Class. ....	20
25	3. Anthropic's Willfulness Will Be Resolved with Common Proof. ....	21
26	4. Damages Will Be Resolved with Common Proof. ....	22
27	B. A Class Action Is Superior to Individual Actions. ....	23
28	In the Alternative, The Court Should Certify the Alternative Classes or Certify an Issue Class Under Rule 23(c)(4). ....	24
	VII. APPOINTMENT OF CLASS COUNSEL .....	24
	VIII. CONCLUSION .....	25

## TABLE OF AUTHORITIES

	Page
<b>Cases</b>	
<i>A&amp;M Records., Inc. v. Napster, Inc.</i> , 239 F.3d 1004 (9th Cir. 2001), as amended (Apr. 3, 2001).....	passim
<i>In re Aimster Copyright Litig.</i> , 334 F.3d 643 (7th Cir. 2003).....	18
<i>Am. Geophysical Union v. Texaco Inc.</i> , 60 F.3d 913 (2d Cir. 1994).....	20
<i>Amchem Prods., Inc. v. Windsor</i> , 521 U.S. 591 (1997).....	10, 14, 15
<i>Amgen Inc. v. Conn. Ret. Plans &amp; Tr. Funds</i> , 568 U.S. 455 (2013).....	10
<i>Aquarian Found., Inc. v. Lowndes</i> , 127 F.4th 814 (9th Cir. 2025).....	15
<i>Bill Graham Archives v. Dorling Kindersley Ltd.</i> , 448 F.3d 605 (2d Cir. 2006).....	21
<i>BMG Music v. Gonzalez</i> , 430 F.3d 888 (7th Cir. 2005).....	18
<i>Briseno v. ConAgra Foods, Inc.</i> , 844 F.3d 1121 (9th Cir. 2017).....	11
<i>Cai v. CMB Exp. LLC</i> , 2024 WL 2334509 (C.D. Cal. Apr. 4, 2024) .....	24
<i>Carnegie v. Household Int’l, Inc.</i> , 376 F.3d 656 (7th Cir. 2004).....	23
 	
<i>In re College Athlete NIL Litig.</i> , 2023 WL 7106483 (N.D. Cal. Sept. 22, 2023) .....	13, 14
<i>Columbia Pictures Indus., Inc. v. Fung</i> , 710 F.3d 1020 (9th Cir. 2013).....	16
<i>Comcast Corp. v. Behrend</i> , 569 U.S. 27 (2013).....	22
<i>In re Conseco Life Ins. Co. LifeTrend Ins. Sales &amp; Mktg. Litig.</i> , 270 F.R.D. 521 (N.D. Cal. 2010).....	13
<i>Danjaq LLC v. Sony Corp.</i> , 263 F.3d 942 (9th Cir. 2001).....	22

**TABLE OF AUTHORITIES**  
(continued)

		Page
1		
2		
3	<i>Ehret v. Uber Techs., Inc.</i> ,	
4	148 F. Supp. 3d 884 (N.D. Cal. 2015) .....	12
5	<div style="background-color: black; height: 1.2em; width: 100%;"></div>	
6	<i>Feist Publ'ns, Inc. v. Rural Tel. Serv. Co.</i> ,	
7	499 U.S. 340 (1991) .....	15
8	<i>Flo &amp; Eddie, Inc. v. Sirius XM Radio, Inc.</i> ,	
9	2015 WL 4776932 (C.D. Cal. May 27, 2015) .....	10, 12, 15, 16
10	<i>Glacier Films (USA), Inc. v. Turchin</i> ,	
11	896 F.3d 1033 (9th Cir. 2018) .....	18
12	<i>Google LLC v. Oracle America, Inc.</i> ,	
13	593 U.S. 1 (2021) .....	17, 19, 21
14	<i>Gulf Oil Co. v. Bernard</i> ,	
15	452 U.S. 89 (1981) .....	10
16	<i>Hachette Book Grp., Inc. v. Internet Archive</i> ,	
17	115 F.4th 163 (2d Cir. 2024) .....	21
18	<i>Harper &amp; Row Publishers, Inc. v. Nation Enters.</i> ,	
19	471 U.S. 539 (1985) .....	17, 20
20	<i>Jimenez v. Allstate Ins. Co.</i> ,	
21	765 F.3d 1161 (9th Cir. 2014) .....	23
22	<i>Just Film, Inc. v. Buono</i> ,	
23	847 F.3d 1108 (9th Cir. 2017) .....	23
24	<i>In re JUUL Labs, Inc., Mktg. Sales Practices &amp; Prods. Liab. Litig.</i> ,	
25	609 F. Supp. 3d 942 (N.D. Cal. 2022) .....	14
26	<i>Keegan v. Am. Honda Motor Co., Inc.</i> ,	
27	284 F.R.D. 504 (C.D. Cal. 2012) .....	14
28	<i>Leyva v. Medline Indus. Inc.</i> ,	
	716 F.3d 510 (9th Cir. 2013) .....	22
	<i>In re Lyft Inc. Sec. Litig.</i> ,	
	2021 WL 3711470 (N.D. Cal. Aug. 20, 2021) .....	11
	<i>Milton H. Greene Archives, Inc. v. Julien's Auction House LLC</i> ,	
	345 F. App'x 244 (9th Cir. 2009) .....	21
	<i>Monge v. Maya Mags., Inc.</i> ,	
	688 F.3d 1164 (9th Cir. 2012) .....	21
	<i>In re Motion to Compel Compliance with Non-Party Subpoenas</i> ,	
	No. 25-mc-80017 (N.D. Cal. Feb. 24, 2025 N.D.C.A.) .....	5, 19

**TABLE OF AUTHORITIES**  
(continued)

	<b>Page</b>
<i>In re Napster, Inc. Copyright Litig.</i> , 2005 WL 1287611 (N.D. Cal. June 1, 2005) .....	<i>passim</i>
<i>Olean Wholesale Grocery Coop., Inc. v. Bumble Bee Foods LLC</i> , 31 F.4th 651 (9th Cir. 2022).....	10
<i>In re Online DVD-Rental Antitrust Litig.</i> , 779 F.3d 934 (9th Cir. 2015).....	13
<i>Parsons v. Ryan</i> , 754 F.3d 657 (9th Cir. 2014).....	11
<i>Perfect 10, Inc. v. Amazon.com, Inc.</i> , 508 F.3d 1146 (9th Cir. 2007).....	19
<i>Ruiz Torres v. Mercer Canyons Inc.</i> , 835 F.3d 1125 (9th Cir. 2016).....	15
<i>Smith v. City of Oakland</i> , 339 F.R.D. 131 (N.D. Cal. 2021) .....	11
<i>Sony Corp. v. Universal City Studios, Inc.</i> , 464 U.S. 417 (1984) .....	21
<i>Thomson Reuters Enter. Ctr. GMBH v. Ross Intel. Inc.</i> , No. 1:20-cv-613, 2025 WL 458520 (D. Del. Feb. 11, 2025).....	3, 20
<i>Triller Fight Club II LLC v. H3 Podcast</i> , 2023 WL 11877604 (C.D. Cal. Sept. 15, 2023).....	19
<i>Tyson Foods, Inc. v. Bouaphakeo</i> , 577 U.S. 442 (2016).....	15
<i>UMG Recordings, Inc. v. MP3.Com, Inc.</i> , 2000 WL 710056 (S.D.N.Y. June 1, 2000).....	17
<i>Valentino v. Carter-Wallace, Inc.</i> , 97 F. 3d 1227 (9th Cir. 1996).....	23, 24
<i>Vaquero v. Ashley Furniture Industries, Inc.</i> , 824 F.3d 1150 (9th Cir. 2016).....	22
<i>Vinole v. Countrywide Home Loans, Inc.</i> , 571 F.3d 935 (9th Cir. 2009).....	24
<i>Wal-Mart v. Dukes</i> , 564 U.S. 338 (2011) .....	11
 <b>Statutes</b>	
17 U.S.C. § 107 .....	17
17 U.S.C. § 410(c) .....	10

**TABLE OF AUTHORITIES**  
(continued)

	<b>Page</b>
17 U.S.C. § 504(b) .....	22
17 U.S.C. § 504(c) .....	22
17 U.S.C. § 504(c)(2) .....	22
17 U.S.C. § 1410(c) .....	15
 <b>Court Rules</b>	
Fed. R. Civ. P. 23(a)(1) .....	11
Fed. R. Civ. P. 23(a)(3) .....	12
Fed. R. Civ. P. 23(a)(4) .....	13
Fed. R. Civ. P. 23(b)(2) .....	13
Fed. R. Civ. P. 23(b)(3) .....	14, 23
Fed. R. Civ. P. 23(g)(1) .....	24
 <b>Treatises</b>	
4 Melville B. Nimmer & David Nimmer, Nimmer on Copyright (Matthew Bender rev. ed. 2016) .....	15, 16



**NOTICE OF MOTION AND MOTION**

TO ALL PARTIES AND THEIR ATTORNEYS OF RECORD:

PLEASE TAKE NOTICE that, on May 15, 2025, at 8:00 a.m., or at some other time as the Court directs, Plaintiffs Andrea Bartz, Andrea Bartz, Inc., Charles Graeber, Kirk Wallace Johnson, and MJ + KJ, Inc. ("Plaintiffs") will move this Court for certification of the following two Classes pursuant to Rule 23 of the Federal Rules of Civil Procedure:

***Pirated Books Class:*** All natural persons, estates, literary trusts, and loan-out companies that are legal or beneficial owners of copyrighted works that: (a) were registered with the United States Copyright Office within five years of the work's publication; (b) were registered with the United States Copyright Office before being downloaded by Anthropic, or within three months of publication; (c) are assigned an International Standard Book Number (ISBN) or Amazon Standard Identification Number (ASIN); and (d) were downloaded by Anthropic as part of [REDACTED].

[REDACTED] ***Books Class:*** All natural persons, estates, literary trusts, and loan-out companies that are legal or beneficial owners of copyrighted works that: (a) were registered with the United States Copyright Office within five years of the work's publication; (b) were registered with the United States Copyright Office within three months of publication or before being [REDACTED]; (c) are assigned an International Standard Book Number (ISBN) or Amazon Standard Identification Number (ASIN); (d) were [REDACTED]; and (e) were used by Anthropic in LLM training.

In the alternative, Plaintiffs request that the Court certify the following dataset-specific classes, which are proposed in lieu of the Pirated Books Class if it is not certified.

**Alternative Class #1 ([REDACTED] Class):** All natural persons, estates, literary trusts, and loan-out companies that are legal or beneficial owners of copyrighted works that: (a) were registered with the United States Copyright Office within five years of the work's publication; (b) were registered with the United States Copyright Office before being downloaded by Anthropic, or within three months of publication; (c) are assigned an International Standard Book Number (ISBN) or Amazon Standard Identification Number (ASIN); and (d) were downloaded by Anthropic as part of [REDACTED].

**Alternative Class #2 ([REDACTED] & [REDACTED] Class):** All natural persons, estates, literary trusts, and loan-out companies that are legal or beneficial owners of copyrighted works that: (a) were registered with the United States Copyright Office within five years of the work's publication; (b) were registered with the United States Copyright Office before being downloaded by Anthropic, or within three months of publication; (c) are assigned an International Standard Book

1 Number (ISBN) or Amazon Standard Identification Number (ASIN); and (d) were  
2 downloaded by Anthropic from [REDACTED] or [REDACTED].<sup>1</sup>

3 Excluded from the foregoing Classes and Alternative Classes are the assigned Judge(s),  
4 Anthropic PBC (“Anthropic”), its officers and directors, members of their immediate families,  
5 their co-conspirators, aiders and abettors, and the heirs, successors or assigns of any of the  
6 foregoing.

7 Plaintiffs also request that they be appointed by the Court as the Class representatives and  
8 that Lieff Cabraser Heimann & Bernstein LLP and Susman Godfrey L.L.P. be appointed as Class  
9 Counsel. The grounds for this motion are that this case meets all the requirements for class  
10 treatment as required under Rules 23(a), 23(b)(2), 23(b)(3), 23(c)(4), and 23(g) of the Federal  
11 Rules of Civil Procedure.

12 The Motion is based on this Notice of Motion and supporting Memorandum of Points and  
13 Authorities; the Declaration of proposed Class Counsel Rachel Geman and Justin A. Nelson; the  
14 Declaration of Jacob S. Miller In Support of Class Certification (“Miller Decl.”) and the exhibits  
15 attached thereto; the expert report of Dr. Ben Yanbin Zhao; the reply briefing in further support of  
16 this Motion; the arguments of counsel; and any such other matters as the Court may consider.

17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  

---

<sup>1</sup> As set out below, *see infra* footnote 2, Plaintiffs continue to discover additional pirated datasets as Anthropic continues to produce discovery.



## I. INTRODUCTION

This case involves Anthropic’s brazen and illegal exploitation of millions of books. For the last several years, Anthropic has needed massive quantities of books to train its large language models (“LLMs”). Instead of buying these books and getting a license to copy them, Anthropic sought out pirated copies from the dark corners of the internet. Anthropic sourced its books from places like [REDACTED] (taken down due to copyright complaints in 2023), [REDACTED] [REDACTED] (shut down by court order multiple times), and [REDACTED] (whose website boasts [REDACTED]). Anthropic built its first models—and the company itself—from the shredded spines and stolen words of pirated and copyrighted books. As was the case for the users of Napster, Kazaa, and other peer-to-peer networks of yesteryear, Anthropic’s decision to download massive quantities of pirated works to avoid paying for or licensing them is an open-and-shut case of largescale copyright infringement. *See A&M Records, Inc. v. Napster, Inc.*, 239 F.3d 1004, 1014–17 (9th Cir. 2001), *as amended* (Apr. 3, 2001).

[REDACTED] *See, e.g., Thomson Reuters Enter. Ctr. GMBH v. Ross Intel. Inc.*, No. 1:20-cv-613, 2025 WL 458520, at \*8–10 (D. Del. Feb. 11, 2025).

Plaintiffs ask the Court to certify two classes alleging claims of copyright infringement: (1) the **Pirated Books Class**, consisting of owners of works that Anthropic downloaded for free from illegal websites and (2) the [REDACTED] **Books Class**, consisting of owners of works that [REDACTED]. Common questions of

1 law and fact will drive the resolution of the copyright claims for both Classes.

2 The claims of all Class members turn on the legality of Anthropic's common course of  
3 conduct. Anthropic did not care about these books on an individualized basis, instead acquiring  
4 books by the tens and hundreds of thousands, compiling them into massive datasets, and using  
5 them as the raw input to train its models. The LLM training process is factory-like: books—  
6 regardless of author, length, genre, or acclaim—are copied, cleaned, copied again, tokenized,  
7 copied again, and fed into the model in those same, standard steps. Anthropic wanted a huge trove  
8 of high-quality "tokens" (*i.e.*, bits of words) that these books as a whole provided because of their  
9 diversity of expressive content.

10 This Court should treat the book authors no differently than Anthropic did—as  
11 commonly-situated. Plaintiffs seek certification of two Classes based on two common courses of  
12 conduct: Anthropic's knowing acquisition of copyright-protected books from pirated sources, and  
13 Anthropic's [REDACTED]. Plaintiffs assert that Anthropic's  
14 unauthorized copying of their works infringed their copyrights. Anthropic, for its part, claims that  
15 its unauthorized pirating or copying of hundreds of thousands of books was fair use. Because  
16 Anthropic acquired books with common methods (downloading [REDACTED]), used  
17 books in a common manner (LLM training), and needed books for a common reason (their  
18 expressive content), its conduct infringed *all* of the copyrights owned by members of each Class,  
19 or *none*.

20 These important issues will be most efficiently resolved by a class action. As Judge Patel  
21 found in the *Napster* case, which involved one technology company's treatment of massive  
22 amounts of copyrighted material in pursuit of profits, "The court thus finds it unlikely that  
23 individual trials of such issues would be practical or efficient." *In re Napster, Inc. Copyright*  
24 *Litig.*, 2005 WL 1287611, at \*12 n.5 (N.D. Cal. June 1, 2005). It would be impractical and  
25 inefficient for individual authors to litigate the same issues again and again, and realistically they  
26 cannot do so. Accordingly, Plaintiffs respectfully ask the Court to certify Copyright Act claims  
27 under Rules 23(b)(2) and 23(b)(3) for the proposed Classes. In the alternative, the Court may  
28 certify certain questions for class treatment.

1 Plaintiffs further ask the Court to appoint Andrea Bartz, Andrea Bartz, Inc., Charles  
 2 Graeber, Kirk Wallace Johnson, and MJ + KJ, Inc. as Class Representatives; and Lieff Cabraser  
 3 Heimann & Bernstein LLP and Susman Godfrey L.L.P. as Class Counsel.

4 **II. COMMON EVIDENCE OF HOW ANTHROPIC VALUES, COPIES, AND USES**  
 5 **PLAINTIFFS' AND CLASS MEMBERS' COPYRIGHTED BOOKS.**<sup>2</sup>

6 **A. Anthropic Benefits from its LLMs, and Needs a High Volume of Books to**  
 7 **Train and Develop Them.**

8 Anthropic was founded in 2021 by former OpenAI employees, chief among them Dario  
 9 Amodei (current Anthropic CEO) and Benjamin Mann. OpenAI has identified Amodei and Mann  
 10 as having been in charge of the “creation and use” of purportedly pirated datasets of books as part  
 11 of OpenAI’s work to train cutting-edge LLMs. *See* Dkt. 17, *In re Motion to Compel Compliance*  
 12 *with Non-Party Subpoenas*, No. 25-mc-80017, at 1–3 (N.D. Cal. Feb. 24, 2025). Here, past was  
 13 prologue. Immediately after its founding, Anthropic’s founders picked up where they left off at  
 14 OpenAI: pirating huge quantities of books from three so-called shadow libraries in order to train  
 15 cutting-edge LLMs. As Amodei said about books: [REDACTED]  
 [REDACTED] Miller Ex. 17.<sup>3</sup>

16 Anthropic used these pirated libraries to create the LLMs that catapulted it to a \$61.5  
 17 billion valuation. *See* Miller Ex. 32. Anthropic makes money directly from—and raised billions  
 18 of dollars from investors for—its already-released “Claude” models, including those trained with  
 19 pirated works. For a monthly fee, businesses and individuals use Claude via Anthropic’s internet  
 20 chatbot or API, a type of software which allows companies to integrate Claude into their own  
 21 products. *See* Anthropic’s Ans. to FAC (Dkt. 72), ¶¶ 1, 5, 21. According to public reports,  
 22 Anthropic has told investors that it projects that annual revenue could reach as high as \$34.5  
 23 billion in 2027. *See* Miller Ex. 33.

24 Anthropic builds its LLMs through a process called training, which consists of feeding a

25 \_\_\_\_\_  
 26 <sup>2</sup> Plaintiffs have summarized facts based on what Anthropic has produced to date in discovery,  
 27 but more than two-thirds of the documents produced thus far were just produced on March 20,  
 one week before this filing, and Plaintiffs also were informed for the first time on March 20 of the  
 existence of additional pirated sources. Miller Decl. ¶ 2; Miller Ex. 1, at 5.

28 <sup>3</sup> “Miller Ex. \_\_\_\_” refers to the exhibits attached to the Declaration of Jacob Miller.

1 massive corpus of text into the model. *See* Anthropic’s Ans. to FAC (Dkt. 72), ¶ 25. The training  
 2 process requires [REDACTED]—a lot of [REDACTED] data. Miller Ex. 7, at 90:5–8; *see* Zhao  
 3 Decl. ¶¶ 68–81. A model trained on small-volume or low-quality data will generate low-quality  
 4 output. *See* Anthropic’s Ans. to FAC (Dkt. 72), ¶¶ 5, 30 (“[T]he phrase ‘garbage in, garbage out’  
 5 is one way to describe that idea.”).

6 More precisely, then, Anthropic needs a large volume of *books* to train its LLMs.  
 7 [REDACTED] Miller Ex. 8. As explained by  
 8 Plaintiffs’ expert Dr. Ben Y. Zhao, Neubauer Professor of Computer Science at University of  
 9 Chicago, books provide LLMs with long-form text that enables them to both process coherent  
 10 long-form text as an input and produce coherent long-form text as an output. *See* Zhao Decl.  
 11 ¶¶ 85–87. Anthropic admits that books are valuable because, *inter alia*, [REDACTED]  
 12 [REDACTED]  
 13 [REDACTED] Miller Ex. 7, at 90:9–91:7. And to Anthropic, *all books* are valuable. In its words,  
 14 [REDACTED]  
 15 [REDACTED]  
 16 Ex. 25, at -37, -46. Notwithstanding books’ undisputed value, *see* Zhao Decl. ¶ 90, Anthropic  
 17 systematically chose to steal books from pirated sources, as shown below, and at no time has paid  
 18 to license any book to train its LLMs.

19 **B. Anthropic Deliberately Copied Books from Pirated Books Datasets.**

20 To acquire books—the high-quality content needed to train its LLMs—Anthropic turned  
 21 to three well-known internet sources for pirated books [REDACTED]  
 22 [REDACTED].

23 [REDACTED]  
 24 [REDACTED]  
 25 [REDACTED]  
 26 [REDACTED]  
 27 [REDACTED]  
 28 [REDACTED]

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28

See Miller Ex. 35.

*Id.*

Three months later, Amodei and others left OpenAI to found Anthropic.

[REDACTED]  
 [REDACTED]  
 [REDACTED]  
 [REDACTED]

Bar Index	Blue Segment (%)	Red Segment (%)	Total Length (%)
1	10	0	10
2	55	45	100
3	90	10	100
4	90	10	100
5	90	10	100
6	85	15	100
7	85	15	100
8	80	20	100
9	80	20	100
10	80	20	100

[REDACTED]

[REDACTED]  
 [REDACTED]  
 [REDACTED]

[REDACTED]  
 [REDACTED]  
 [REDACTED]

\_\_\_\_\_

\_\_\_\_\_

As with its piracy, Anthropic launched [REDACTED] to keep up with its primary competitor, OpenAI. Anthropic told investors:

[REDACTED]  
 [REDACTED]  
 [REDACTED]  
 [REDACTED]



**D. Anthropic Copied Books to Create Training Datasets and Train and Develop its LLMs.**

After acquiring Plaintiffs' and Class members' copyrighted books through the common methods identified above, Anthropic copied, stored, and used books for its benefit.

*First*, for those books used as pre-training data, Anthropic made at least three copies:

See Miller Ex. 1, at 6–8; Zhao Decl. ¶ 74–75.

Plaintiffs' inspection of Anthropic's datasets produced in discovery indicates that Anthropic's [REDACTED] dataset contains [REDACTED] books, that its [REDACTED] datasets contain up to an estimated [REDACTED] books, that its [REDACTED] datasets contain an estimated [REDACTED] books, and that its [REDACTED] datasets contain (depending on the set in question) an estimated [REDACTED] books respectively. See Zhao Decl. ¶¶ 18–41.

See Zhao Decl. ¶¶ 42–68; Miller Ex. 4, at 2–7; Ex. 7.

*Second*, Anthropic's model training process consists of standard steps, including [REDACTED], all of which are common across Anthropic's training datasets. See Miller Ex. 1, at 6–8; Zhao Decl. ¶¶ 69–78.

**E. Anthropic's Use of Pirated and [REDACTED] Books Affected the Book-Purchase Market and the LLM-Licensing Market.**

Anthropic's common conduct adversely impacted book-purchasing and licensing for LLM usage for Class members in a common way. For example, common evidence of comparable licenses will show not only that there is a licensing market for books, but that intermediaries exist for efficiency. As one example, an LLM company recently contracted with the publisher Harper Collins to pay a flat fee of \$5,000 per book for a license to use the book in model training for three years. See Miller Ex. 39. Common evidence will show that license fees are artificially suppressed because Anthropic and its peers have downloaded and used pirated books for free. But

1 fair-market license fees will be established on common evidence and expert testimony about the  
2 licensing market.

3 **F. Plaintiffs Are Copyright Holders Who Bring a Common Copyright Act**  
4 **Claim.**

5 Plaintiffs, which include both authors who make their living by writing and their loan-out  
6 companies, have written and/or own the copyright of books registered with the U.S. Copyright  
7 Office. *See* Miller Ex. 30 (named plaintiff registrations). They each bring claims for copyright  
8 infringement against Anthropic arising out of Anthropic’s piracy of their works and the copies  
9 Anthropic made when it used their works to train its LLMs. Each asserted copyrighted work has  
10 been registered with the United States Copyright Office within five years of publication, which  
11 establishes “prima facie evidence of the validity of the copyright.” 17 U.S.C. § 410(c).

12 **III. LEGAL STANDARD**

13 “Class actions serve an important function in our system of civil justice.” *Gulf Oil Co. v.*  
14 *Bernard*, 452 U.S. 89, 99 (1981). Rule 23’s purpose is to provide for the efficient administration  
15 of justice, as the class action mechanism allows large numbers of claims involving the same core  
16 issues to proceed in the aggregate, providing a path to relief where otherwise there is none. The  
17 purpose of a ruling on a class certification motion “is to select the ‘method’ best suited to  
18 adjudication of the controversy ‘fairly and efficiently.’” *Amgen Inc. v. Conn. Ret. Plans & Tr.*  
19 *Funds*, 568 U.S. 455, 460 (2013). Class certification is proper if Plaintiffs satisfy the requirements  
20 of Rule 23(a) and one of the prongs of Rule 23(b). *Olean Wholesale Grocery Coop., Inc. v.*  
21 *Bumble Bee Foods LLC*, 31 F.4th 651, 663 (9th Cir. 2022) (en banc). Cases like this one, arising  
22 from a single course of conduct that affects large numbers of plaintiffs, are particularly amenable  
23 to class treatment. *See, e.g., Amchem Prods., Inc. v. Windsor*, 521 U.S. 591, 625 (1997).

24 Applying these standards, courts in and out the Ninth Circuit have certified copyright  
25 classes involving allegations of largescale infringement arising out of a common course of  
26 conduct. *See, e.g., In re Napster*, 2005 WL 1287611, at \*12; *Flo & Eddie, Inc. v. Sirius XM*  
27 *Radio, Inc.*, 2015 WL 4776932, at \*17 (C.D. Cal. May 27, 2015).

1 **IV. THE PROPOSED CLASSES SATISFY RULE 23(a).**

2 **A. The Proposed Classes Are Objectively Defined.**

3 The Ninth Circuit has not adopted any ascertainability requirement. *Briseno v. ConAgra*  
 4 *Foods, Inc.*, 844 F.3d 1121, 1124–25 n.4 (9th Cir. 2017). Here, both the Pirated Books Class and  
 5 the [REDACTED] Books Class definitions nonetheless rely on objective criteria verifiable through:  
 6 (1) official records showing each Class member held a copyright “registered with the United  
 7 States Copyright Office” (and the registration date), and (2) Anthropic’s records identifying the  
 8 books illegally [REDACTED] downloaded, and/or [REDACTED], copied, and used to train and develop its  
 9 LLMs. *See* Miller Ex. 30 (named plaintiff registrations); Zhao Decl. ¶¶ 42–68.

10 **B. The Proposed Classes Are Sufficiently Numerous.**

11 The proposed Classes meet Rule 23(a)’s numerosity requirement. Before a class can be  
 12 certified, the court must determine that it is “so numerous that joinder of all members is  
 13 impracticable.” *See* Fed. R. Civ. P. 23(a)(1). Courts generally find “a presumption of  
 14 impracticability of joinder” for classes above 40 members. *See, e.g., Smith v. City of Oakland*,  
 15 339 F.R.D. 131, 138 (N.D. Cal. 2021) (cleaned up). Numerosity is easily satisfied here because  
 16 the (at least) hundreds of thousands of books in each proposed class, *see* Zhao Decl. ¶¶ 18–41,  
 17 implicate the rights of hundreds of thousands or more copyright holders.

18 **C. Common Issues of Fact and Law Exist.**

19 The proposed Classes also satisfy Rule 23(a)(2)’s commonality requirement. “A  
 20 contention is sufficiently common where it is capable of class-wide resolution—which means that  
 21 determination of its truth or falsity will resolve an issue that is central to the validity of each one  
 22 of the claims in one stroke.” *In re Lyft Inc. Sec. Litig.*, 2021 WL 3711470, at \*3 (N.D. Cal. Aug.  
 23 20, 2021) (cleaned up). “[E]ven a single common question” will do. *Parsons v. Ryan*, 754 F.3d  
 24 657, 675 (9th Cir. 2014) (quoting *Wal-Mart v. Dukes*, 564 U.S. 338, 359 (2011)).

25 Common issues here include whether and how Anthropic obtained books; whether and  
 26 how Anthropic copied and stored books; the nature of Anthropic’s use of books in its LLMs;  
 27 whether Anthropic’s conduct is fair use; whether Anthropic’s conduct was willful; whether  
 28 Anthropic’s uses were commercial; whether there are markets or potential markets for book

1 purchasing and LLM data licensing; whether Anthropic made profits; and the remedies to which  
 2 Plaintiffs and the Classes are entitled.

3 Anthropic didn't treat books any differently from one another: it acquired them in droves  
 4 from common sources, cleaned and processed them in a standardized manner, and fed them into  
 5 its models as a collective input. Miller Ex. 1, at 6–8; Zhao Decl. ¶¶ 69-78. That common factual  
 6 predicate is more than enough to satisfy the commonality requirement. *See In re Napster*, 2005  
 7 WL 1287611, at \*3 (finding commonality where “the claims of each member of the class  
 8 fundamentally arise from the same factual predicate: namely, that Napster users infringed the  
 9 class member’s exclusive rights in one or more copyrighted musical compositions”); *Flo &*  
 10 *Eddie*, 2015 WL 4776932, at \*9 (finding common questions “that the Court can address and  
 11 resolve as to the entire class members in ‘one stroke’ because class members argue that Sirius  
 12 XM engaged in the same exploitative conduct against them (broadcasting and streaming their pre-  
 13 1972 recording without authorization).”).

14 **D. Plaintiffs’ Claims Are Typical of the Classes’ Claims.**

15 Plaintiffs’ claims satisfy Rule 23(a)(3)’s typicality requirement. This rule requires  
 16 Plaintiffs to show that the “claims or defenses of the representative parties are typical of the  
 17 claims or defenses of the class.” Fed. R. Civ. P. 23(a)(3). Typicality is satisfied “when each class  
 18 member’s claim arises from the same course of events, and each class member makes similar  
 19 legal arguments to prove the defendant’s liability.” *Ehret v. Uber Techs., Inc.*, 148 F. Supp. 3d  
 20 884, 892 (N.D. Cal. 2015) (cleaned up).

21 Here, Plaintiffs’ claims are identical to the claims of all members of the Pirated Books  
 22 Class and [REDACTED] Books Class because they arise from the exact same course of conduct and  
 23 under the same legal claim. Each Plaintiff (either individually or through their Plaintiff loan-out  
 24 company) owns the copyright to: (1) a book Anthropic [REDACTED] or downloaded from a shadow  
 25 library of pirated material without paying a purchase price, and (2) a book Anthropic [REDACTED]  
 26 [REDACTED]. *See* Zhao Decl. ¶¶ 42–68; Miller Ex. 7; Ex. 30 (named plaintiff registrations).  
 27 Because Plaintiffs’ works were subject to the same infringement impacting all other Class  
 28 members, Plaintiffs satisfy the typicality requirement.

1           **E. Plaintiffs Fairly and Adequately Represent the Classes.**

2           Plaintiffs also meet Rule 23(a)(4)’s adequacy requirement. Rule 23(a)(4) requires that “the  
3           representative parties will fairly and adequately protect the interests of the class.” Fed. R. Civ. P.  
4           23(a)(4). The adequacy requirement examines whether the proposed representatives have any  
5           conflicts of interests with the proposed Classes and that Plaintiffs are represented by qualified and  
6           competent counsel. *In re Conseco Life Ins. Co. LifeTrend Ins. Sales & Mktg. Litig.*, 270 F.R.D.  
7           521, 531 (N.D. Cal. 2010). Only conflicts that are “fundamental to the suit and that go to the heart  
8           of the litigation” are relevant. *In re Online DVD-Rental Antitrust Litig.*, 779 F.3d 934, 942 (9th  
9           Cir. 2015) (cleaned up).

10           Here, Plaintiffs’ interests are aligned with those of both Classes because Plaintiffs and  
11           Class members have been harmed by—and have an interest in maximizing recovery in and  
12           compensation for—Anthropic’s largescale copying of their books, including to train its LLMs.  
13           Additionally, Plaintiffs are engaged participants in this litigation. They have stayed abreast of  
14           case developments, responded to discovery requests, testified at depositions, and made clear that  
15           they are willing to testify at trial. *See* Miller Ex. 27, at 49:16–52:25, 54:23–55:4, 84:22–85:25;  
16           Ex. 28, at 43:19–22, 44:10–12, 75:24–76:8; Ex. 29, at 21:9–18, 60:3–61:2, 61:17–21, 63:18–24.  
17           And Plaintiffs are represented by qualified and competent counsel. *See supra* § VII. Accordingly,  
18           the Court should appoint Plaintiffs as Class Representatives.

19           **V. THE PROPOSED CLASSES SATISFY RULE 23(b)(2).**

20           Class certification is appropriate under Rule 23(b)(2) where, as here, “the party opposing  
21           the class has acted or refused to act on grounds that apply generally to the class, so that final  
22           injunctive relief or corresponding declaratory relief is appropriate respecting the class as a  
23           whole.” Fed. R. Civ. P. 23(b)(2). “[T]he Rule 23(b)(2) requirements are unquestionably satisfied  
24           when members of a putative class seek uniform injunctive or declaratory relief from policies or  
25           practices that are generally applicable to the class as a whole.” *In re College Athlete NIL Litig.*,  
26           2023 WL 7106483, at \*6 (N.D. Cal. Sept. 22, 2023) (cleaned up). “That inquiry does not require  
27           an examination of the viability or bases of the class members’ claims for relief, does not require  
28           that the issues common to the class satisfy a Rule 23(b)(3)-like predominance test, and does not



1 require a finding that all members of the class have suffered identical injuries.” *Id.* “Rather, as the  
 2 text of the rule makes clear, this inquiry asks only whether the party opposing the class has acted  
 3 or refused to act on grounds that apply generally to the class.” *Id.* (cleaned up).

4 Here, Anthropic’s course of conduct makes final injunctive relief appropriate. If Plaintiffs  
 5 prevail on their copyright claim, injunctive relief precluding Anthropic from copying or using the  
 6 Class Works to train its LLMs without a license would inure to the benefit of all Class members.  
 7 With respect to both the Pirated Books Class and [REDACTED] Books Class, Anthropic’s conduct has  
 8 “applied in the past, in a uniform manner to all members,” and Plaintiffs’ requested relief “would  
 9 enjoin the enforcement of the challenged [conduct],” thereby “provid[ing] uniform relief to all  
 10 members of the proposed class.” *In re College Athlete*, 2023 WL 7106483, at \*6.

#### 11 **VI. THE PROPOSED CLASSES SATISFY RULE 23(b)(3).**

12 To obtain certification of a class under Rule 23(b)(3), Plaintiffs must show that “the  
 13 questions of law or fact common to class members predominate over any questions affecting only  
 14 individual members, and that a class action is superior to other available methods for fairly and  
 15 efficiently adjudicating the controversy.” Fed. R. Civ. P. 23(b)(3). Here, common questions of  
 16 law and fact predominate over individual ones and the class action mechanism is superior to other  
 17 methods of adjudication.

##### 18 **A. Common Questions of Law and Fact Predominate.**

19 Rule 23(b)(3) considers whether questions common to the class predominate. The  
 20 predominance requirement “tests whether proposed classes are sufficiently cohesive to warrant  
 21 adjudication by representation.” *Amchem*, 521 U.S. at. 623. “If common questions present a  
 22 significant aspect of the case and they can be resolved for all members of the class in a single  
 23 adjudication, then there is clear justification for handling the dispute on a representative rather  
 24 than on an individual basis, and the predominance test is satisfied.” *Keegan v. Am. Honda Motor*  
 25 *Co., Inc.*, 284 F.R.D. 504, 526 (C.D. Cal. 2012) (cleaned up). “Predominance is not a counting  
 26 game.” *In re JUUL Labs, Inc., Mktg. Sales Practices & Prods. Liab. Litig.*, 609 F. Supp. 3d 942,  
 27 967 (N.D. Cal. 2022). “Rather, more important questions apt to drive the resolution of the  
 28 litigation are given more weight in the predominance analysis over individualized questions



1 which are of considerably less significance to the claims of the class.” *Ruiz Torres v. Mercer*  
 2 *Canyons Inc.*, 835 F.3d 1125, 1134 (9th Cir. 2016).<sup>4</sup>

3 Predominance is “readily met” where, as here, the primary factual and legal questions  
 4 center on the defendant’s course of conduct. *See Amchem*, 521 U.S. at 625. The core issues that  
 5 will drive resolution of Class members’ claims—including whether Anthropic copied Pirated  
 6 Books and [REDACTED] Books, whether Anthropic’s copying is fair use, what damages should be  
 7 awarded, and whether Anthropic’s conduct was willful—turn entirely on Anthropic’s uniform  
 8 course of conduct. Because each of these issues presents common questions with common  
 9 answers, common issues of fact and law predominate here.

10 **1. Plaintiffs’ Prima Facie Case of Copyright Infringement Will Rise or**  
 11 **Fall on Common Proof.**

12 Copyright infringement consists of only two elements: “(1) ownership of a valid  
 13 copyright, and (2) copying of constituent elements of the work that are original.” *Feist Publ’ns,*  
 14 *Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 361 (1991). Copyright infringement is a strict liability  
 15 offense. *See, e.g.*, 4 Melville B. Nimmer & David Nimmer, *Nimmer on Copyright*, § 13.08(C)(1)  
 16 (Matthew Bender rev. ed. 2016).

17 *First*, each member of each proposed Class owns the copyrights to a Class Work  
 18 registered within five years of publication. Miller Ex. 30 (named plaintiff registrations).  
 19 Ownership can thus be demonstrated through registration records from the Copyright Office. 17  
 20 U.S.C. § 1410(c); *Aquarian Found., Inc. v. Lowndes*, 127 F.4th 814, 819 (9th Cir. 2025)  
 21 (affirming reliance on registrations as prima facie evidence of plaintiff’s authorship). Courts find  
 22 predominance in copyright class actions even where questions of ownership may be subject to  
 23 individualized issues. *See, e.g., In re Napster*, 2005 WL 1287611, at \*7 (predominance satisfied  
 24 even though “it is true that proof of ownership, registration, and actual damages ultimately  
 25 requires a work-by-work inquiry”). Consider *Flo & Eddie*, in which the court certified a class

26 \_\_\_\_\_  
 27 <sup>4</sup> “[W]hen one or more of the central issues in the action are common to the class . . . , the action  
 28 may be considered proper under Rule 23(b)(3) even though other important matters will have to  
 be tried separately, such as damages or some affirmative defenses peculiar to some individual  
 class members.” *Tyson Foods, Inc. v. Bouaphakeo*, 577 U.S. 442, 453 (2016) (cleaned up).

1 bringing copyright infringement claims on behalf of owners of pre-1972 sound recordings  
 2 because, although “[o]wnership of the subject recordings. . . is an individual issue,” the process  
 3 could be streamlined by affidavit such that individual questions would not “predominate common  
 4 issues of fact and law.”<sup>5</sup> 2015 WL 4776932, at \*10–12.

5 *Second*, Anthropic’s downloading of copyrighted material “constitute[s] direct  
 6 infringement” of a copyright holder’s exclusive right of reproduction. *See Napster*, 239 F.3d at  
 7 1014; *Columbia Pictures Indus., Inc. v. Fung*, 710 F.3d 1020, 1034 (9th Cir. 2013)  
 8 (“downloading copyrighted material . . . violates the copyright holder’s . . . right to  
 9 reproduction”); *see also* 4 Nimmer on Copyright § 13D.02 (“For instance, photocopying the  
 10 entirety of a novel is plainly actionable.”). Anthropic made multiple, unlicensed copies of the  
 11 Pirated Books and [REDACTED] Books through its standard training process. *See Zhao Decl.* ¶¶ 69–  
 12 78. Anthropic’s standard training process is common evidence.

13 Common evidence will prove that Anthropic downloaded and/or trained on unlicensed  
 14 copies of the Pirated Books and [REDACTED] Books. *See Zhao Decl.* ¶¶ 42–68. [REDACTED]  
 15 [REDACTED]  
 16 [REDACTED] Miller Ex. 6, at 38:19–39:2; Ex. 7. Common evidence from these datasets (or  
 17 materials about these datasets) will show precisely which books are at issue. Specifically,  
 18 Plaintiffs’ expert has shown how algorithms can be used to precisely verify that an individual  
 19 book is in a given dataset. *See Zhao Decl.* ¶¶ 53–67. The registration materials discussed above,  
 20 the metadata which Anthropic has produced about its datasets, and the datasets themselves  
 21 provide a common basis to establish Anthropic’s liability for copyright infringement. *In re*  
 22 *Napster*, 2005 WL 1287611, at \*7 (finding predominance because “the claims of every member  
 23 of the class are uniformly premised upon the uploading or downloading of a copyrighted work by  
 24 Napster users [and defendant’s] involvement in the operation of the Napster network”). There is  
 25 no reason for this case to proceed as thousands (or more) individual adjudications when the

26 \_\_\_\_\_  
 27 <sup>5</sup> This case is more straightforward than *Flo & Eddie*, which involved pre-1972 sound recordings,  
 28 including unregistered works with more complex ownership determinations. *Id.* at \*7.

1 common proof establishing liability is a handful of discrete datasets.

## 2 **2. Fair Use Will Be Resolved with Common Proof.**

3 Anthropic's fair use affirmative defense will also be resolved with common evidence that  
4 presents common issues of fact and law, including (with respect to the Pirated Books Class) the  
5 antecedent question of whether fair use as a defense is available at all. The class mechanism will  
6 help to answer these questions on a consistent basis for hundreds of thousands of common claims.

7 The affirmative defense of fair use presents a "mixed question of law and fact." *Google*  
8 *LLC v. Oracle America, Inc.*, 593 U.S. 1, 24 (2021) (quoting *Harper & Row Publishers, Inc. v.*  
9 *Nation Enters.*, 471 U.S. 539, 560 (1985)). The Copyright Act lists four nonexclusive factors for  
10 evaluating whether otherwise infringing conduct is fair use:

11 (1) the purpose and character of the use, including whether such use is of a  
12 commercial nature or is for nonprofit educational purposes;

13 (2) the nature of the copyrighted work;

14 (3) the amount and substantiality of the portion used in relation to the copyrighted  
work as a whole; and

15 (4) the effect of the use upon the potential market for or value of the copyrighted  
work.

16 17 U.S.C. § 107.

17 As shown below, whether Anthropic's brazen copying of the Pirated Books and t  
18   
19 Books without a license focuses exclusively on Anthropic's conduct—and the common  
20 defense of fair use (including its availability in the first instance)—can be answered on a class-  
21 wide basis using common evidence.

### 22 **a. Pirated Books Class.**

23 Plaintiffs allege that Anthropic—like Napster users of yesteryear—committed copyright  
24 infringement by   
25 and/or downloading "for free something they would ordinarily have to  
26 buy" to compile the three pirated datasets revealed thus far in discovery. *Napster*, 239 F.3d at  
27 1015; *see also UMG Recordings, Inc. v. MP3.Com, Inc.*, 2000 WL 710056, at \*1 (S.D.N.Y. June  
28 1, 2000) (The "mere fact" that copyright infringement is "clothed in the exotic webbing of" a new  
technology "does not disguise its illegality."). As a threshold matter, there is a (common) question

1 about whether Anthropic can even assert fair use with respect to the Pirated Books Class.  
 2 Plaintiffs submit that the downloading of copyrighted material without permission is itself illegal  
 3 and effectively ends the fair use analysis. Plaintiffs are aware of no case—in any jurisdiction at  
 4 any time—where a court found that the *initial illegal download* was fair use due to the use to  
 5 which the illegal downloaded material was subsequently put. *See, e.g., In re Aimster Copyright*  
 6 *Litig.*, 334 F.3d 643, 645 (7th Cir. 2003) (Posner, J.) (affirming preliminary injunction against  
 7 file-sharing website on contributory infringement theory, because Aimster’s users were the  
 8 “direct infringers” for downloading music from Aimster’s site without paying).

9 Further, any analysis will entail reliance on common evidence.

10 **First Factor:** Common proof confirms the nature of Anthropic’s use—that is, that  
 11 Anthropic downloaded [REDACTED] approximately several million books wholesale. Zhao  
 12 Decl. ¶¶ 23–34. Courts have consistently held that downloading bootleg electronic copies of  
 13 copyrighted works even to avoid paying for *personal, noncommercial* use is *nontransformative*.  
 14 *See, e.g., Napster*, 239 F.3d at 1015 (“exploitative unauthorized cop[ying] of copyrighted works”  
 15 in order “to save the expense of purchasing authorized copies” is not fair use); *BMG Music v.*  
 16 *Gonzalez*, 430 F.3d 888, 891 (7th Cir. 2005) (“downloading full copies of copyrighted material  
 17 [via P2P network] without compensation to authors cannot be deemed ‘fair use’” as “copiers such  
 18 as [defendant] cannot ask courts (and juries) to second-guess the market and call wholesale  
 19 copying ‘fair use’”); *Glacier Films (USA), Inc. v. Turchin*, 896 F.3d 1033, 1043 (9th Cir. 2018)  
 20 (dismissing as “baseless” the argument that “downloading or uploading of the copyrighted work”  
 21 via a P2P network “was permitted by the doctrine of fair use”). Thus, as in these many other  
 22 Copyright Act cases, the Court may determine that “downloading [] files does not transform the  
 23 copyrighted work.” *See, e.g., Napster*, 239 F.3d at 1015 (upholding denial of fair use).

24 In addition, common evidence confirms that Anthropic downloaded these works for a  
 25 commercial purpose. Anthropic’s internal records confirm that Anthropic used Pirated Books  
 26 from the [REDACTED] dataset to train and develop its LLMs. Miller Ex. 7 ([REDACTED])  
 27 [REDACTED]. [REDACTED]  
 28 [REDACTED]



Whether

Anthropic's admitted downloading, [REDACTED] and use of these datasets—for training, research, and development for the benefit of a for-profit company—are “commercial” is a common question.

**Second and Third Factors:** The second and third factors rise or fall on common proof, too. As to the nature of the copyrighted work, each of the copyrighted works are books, which are at the core of copyright protection under the statute. *See Google*, 593 U.S. at 23. As to the amount and substantiality of the portion used, Anthropic's datasets, including those processed for training, provide common proof of how much of each work Anthropic acquired via downloading and whether Anthropic used the text of each copyrighted work in training. *See Napster*, 239 F.3d at 1016 (“file transfer necessarily ‘involves copying the entirety of the copyrighted work’”).

**Fourth Factor:** This analysis will likewise rise or fall with common proof. Here, there is common evidence that Anthropic relied on plainly pirated sources to [REDACTED] and download pirated books for free. This is relevant to whether Anthropic's conduct harmed the traditional book-sale market for Plaintiffs' and Class members' books, as well as the developing market for licensing books for LLM training, because instead of paying a purchase price or licensing fee, Anthropic sought free versions online. *See Napster*, 239 F.3d at 1015 (unauthorized copying in order “to save the expense of purchasing authorized copies” is not fair use).

Finally, Anthropic's incorrect assertion that its internet piracy is irrelevant to the fair use analysis because it relies on considerations of Anthropic's “bad faith” is yet another legal question common to the class. *Compare* Miller Ex. 31, at 11:20–12:25 (Anthropic arguing at the Feb. 25 hearing that bad faith is not relevant here) *with Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146, 1164 n.8 (9th Cir. 2007) (“the general rule [is] that a party claiming fair use must act in a manner generally compatible with principles of good faith and fair dealing”); *Triller Fight Club II LLC v. H3 Podcast*, 2023 WL 11877604, at \*8 (C.D. Cal. Sept. 15, 2023) (analyzing bad faith in fair use defense because “[u]ntil and unless the Supreme Court rules that the defendants’

1 good faith—or lack thereof—is not a relevant consideration, this Court is bound by the holdings  
2 of the Ninth Circuit.”).

3 Anthropic downloaded millions of books in at least three large tranches for a commercial  
4 purpose and without paying a purchase price or licensing fee. Whether those free downloads are  
5 fair use, or not, presents a common question with a common answer.

6 **b. [REDACTED] Books Class.**

7 Whether the fair use defense applies to Anthropic’s use of [REDACTED] Books to train its  
8 LLMs is also a common question with a common answer.

9 **First Factor:** Whether Anthropic’s acquisition and use of [REDACTED] books to develop its  
10 Claude models was commercial is a common question. Anthropic “stands to profit from  
11 exploitation of the copyrighted material without paying” on a common basis. *Harper & Row*, 471  
12 U.S. at 562. [REDACTED]

13 [REDACTED]  
14 [REDACTED] Anthropic’s own  
15 documents will prove Plaintiffs’ claim that Anthropic’s use was commercial, including by  
16 showing the financial calculus and competitive motive behind those decisions. *See, e.g.*, Miller  
17 Ex. 8.

18 Anthropic foreseeably will claim that its use of books in training its models is  
19 transformative, but this too raises common questions with common answers. Plaintiffs maintain  
20 that the LLM training process—in which Anthropic makes a series of copies of each book—is  
21 nontransformative because it is designed to mine and exploit the expressive elements of each  
22 work. *See* Miller Ex. 13, at -109 ([REDACTED]  
23 [REDACTED]); *see also Am. Geophysical Union v. Texaco Inc.*, 60 F.3d 913, 916 (2d Cir.  
24 1994) (holding that Texaco’s “institutional, systematic copying [to] increase[] the number of  
25 copies available to scientists while avoiding the necessity of paying for license fees or for  
26 additional subscriptions” was not fair use); *Ross Intel.*, 2025 WL 458520, at \*8–10 (holding that  
27 the reproduction of Westlaw headnotes to train an AI model was not fair use). Anthropic, relying  
28 on the “intermediate copying” cases, disagrees, but whether Anthropic’s uniform training process



1 is nontransformative exploitation or technically necessary intermediate copying is a common  
2 question with a common answer.

3 ***Second and Third Factors:*** The second and third factors for the [REDACTED] Books Class  
4 will be exactly the same as for the Pirated Books Class and will therefore rise or fall on common  
5 proof. Anthropic [REDACTED] and  
6 used them in a standardized way: to train its LLMs.

7 ***Fourth Factor:*** For this Class, this factor will turn on whether or not there is a potential  
8 licensing market to train LLMs on copyrighted books, another common question with a common  
9 answer. In evaluating the fourth factor, courts determine “whether there was harm to the actual or  
10 potential markets for the copyrighted work.” *Google*, 593 U.S. at 24. Such markets may be  
11 immediate, delayed, or future, and include a market for derivative works. *Monge v. Maya Mags.,*  
12 *Inc.*, 688 F.3d 1164, 1181 (9th Cir. 2012). “It is indisputable that, as a general matter, a copyright  
13 holder is entitled to demand a royalty for licensing others to use its copyrighted work, and that the  
14 impact on potential licensing revenues is a proper subject for consideration in assessing the fourth  
15 factor.” *Hachette Book Grp., Inc. v. Internet Archive*, 115 F.4th 163, 192 (2d Cir. 2024)  
16 (quoting *Bill Graham Archives v. Dorling Kindersley Ltd.*, 448 F.3d 605, 614 (2d Cir. 2006)). The  
17 key common question here is whether there *is* a potential market for the use of books as training  
18 data, which the parties will dispute, with Anthropic likely claiming there is no market categorically  
19 and Plaintiffs demonstrating that indeed there is (as evident in agreements already reached). *See*  
20 *Miller Ex. 39*. Class treatment is warranted. Notably, another common question under the fourth  
21 factor is whether Anthropic’s common usage gives rise to a class-wide inference or presumption of  
22 market harm. “‘If the intended use is for commercial gain, that likelihood of market harm may be  
23 presumed.’” *Napster*, 239 F.3d at 1016 (emphasis in original) (quoting *Sony Corp. v. Universal*  
24 *City Studios, Inc.*, 464 U.S. 417, 451 (1984)). Thus, common issues predominate as to the fair use  
25 analysis for the [REDACTED] Books Class.

### 26 **3. Anthropic’s Willfulness Will Be Resolved with Common Proof.**

27 In the context of copyright infringement, conduct is willful if it occurs “with knowledge  
28 that the defendant’s conduct constitutes copyright infringement.” *Milton H. Greene Archives, Inc.*

1 *v. Julien’s Auction House LLC*, 345 F. App’x 244, 248 (9th Cir. 2009) (quoting *Danjaq LLC v.*  
 2 *Sony Corp.*, 263 F.3d 942, 957 (9th Cir. 2001)). If infringement is willful, the range of statutory  
 3 damages increases. 17 U.S.C. § 504(c)(2). Anthropic’s state of mind presents yet another  
 4 common issue. *See In re Napster*, 2005 WL 1287611, at \*4 (“far from counseling  
 5 against class treatment of plaintiffs’ claims, the ‘knowledge’ and ‘willfulness’ inquiries give rise  
 6 to a number of common issues of law and fact that support certification of the plaintiff class.”).  
 7 Anthropic’s knowledge cannot possibly vary by Class member. Thus, the central question of  
 8 whether Anthropic’s conduct was willful is common, would be most efficiently answered on a  
 9 class-wide basis, and further establishes that common issues predominate.

#### 10 **4. Damages Will Be Resolved with Common Proof.**

11 At the class certification stage, the predominance inquiry looks to whether “damages are  
 12 susceptible of measurement across the entire class.” *Comcast Corp. v. Behrend*, 569 U.S. 27, 35  
 13 (2013). “[T]he presence of individualized damages cannot, by itself, defeat class certification  
 14 under Rule 23(b)(3).” *Leyva v. Medline Indus. Inc.*, 716 F.3d 510, 514 (9th Cir. 2013).

15 Here, Plaintiffs have two measures of damages, both of which are susceptible to class-  
 16 wide measurement: statutory damages, or actual damages plus disgorgement of profits. If  
 17 Plaintiffs seek statutory damages, damages are basic arithmetic: multiplying the number of  
 18 copyrighted books Anthropic used by the statutory penalty. *See* 17 U.S.C. § 504(c). *See, e.g., In*  
 19 *re Napster*, 2005 WL 1287611, at \*10 (certifying class of copyright holders that had yet to elect  
 20 to seek statutory or actual damages).

21 Alternatively, Plaintiffs could seek disgorgement of Anthropic’s profits obtained from  
 22 copyright infringement. *See* 17 U.S.C. § 504(b). Disgorgement is a collective remedy that would  
 23 be determined by common evidence of Anthropic’s own finances.

24 With respect to actual damages, the common evidence shows that actual damages do *not*  
 25 vary book-to-book, as shown by the Harper Collins licensing deal for a flat \$5,000 per book. *See*  
 26 *Miller Ex. 39*. And even if Anthropic is able to establish some degree of variation with respect to  
 27 individual actual damages estimates, and Plaintiffs elect to pursue actual damages, “the need for  
 28 individual damages calculations does not, alone, defeat class certification.” *Vaquero v. Ashley*

*Furniture Industries, Inc.*, 824 F.3d 1150, 1155 (9th Cir. 2016); *see also Jimenez v. Allstate Ins. Co.*, 765 F.3d 1161, 1168 (9th Cir. 2014) (“So long as the plaintiffs were harmed by the same conduct, disparities in how or by how much they were harmed [does] not defeat class certification.”).

**B. A Class Action Is Superior to Individual Actions.**

Rule 23(b)(3) also requires that “a class action is superior to other available methods for fairly and efficiently adjudicating the controversy.” Fed. R. Civ. P. 23(b)(3). Courts consider:

(A) the class members’ interests in individually controlling the prosecution or defense of separate actions; (B) the extent and nature of any litigation concerning the controversy already begun by or against class members; (C) the desirability or undesirability of concentrating the litigation of the claims in the particular forum; and (D) the likely difficulties in managing a class action.

Fed. R. Civ. P. 23(b)(3). A class action is a superior means of adjudicating a dispute if class litigation of common issues “reduce[s] litigation costs and promote[s] greater efficiency.”

*Valentino v. Carter-Wallace, Inc.*, 97 F. 3d 1227, 1234–35 (9th Cir. 1996).

A class action is the superior method of adjudication here for several reasons. Few if any Class members could afford on their own to bring an expensive copyright case against a well-funded artificial intelligence company with extensive litigation resources, especially when compared to the available damages for any one copyright owner. While not insignificant, the damages available under the Copyright Act are insufficient to compensate the average rights holder for the time and cost required to bring an action against a multi-billion-dollar corporation. Beyond the expense of attorneys’ fees, the expert work involved in developing the evidence necessary to succeed on these claims is likely to be cost-prohibitive. Inspecting Anthropic’s training data and source code alone takes days of expert work. *See Just Film, Inc. v. Buono*, 847 F.3d 1108 (9th Cir. 2017) (finding that a class action was superior in part because of the “likely need for expert testimony”). A class action is therefore the superior—indeed, the only practical—method of resolving these claims. *See Carnegie v. Household Int’l, Inc.*, 376 F.3d 656, 661 (7th Cir. 2004) (noting that “[t]he *realistic* alternative to a class action is not 17 million individual suits, but zero individual suits”) (emphasis in original). The obvious benefit of class treatment, therefore, is that it is the only means by which these claims will be heard.

Furthermore, class treatment reduces the burden on courts, which would otherwise have to manage and resolve repeated, individual actions, creating the risk of duplicative trials and inconsistent rulings. *See Vinole v. Countrywide Home Loans, Inc.*, 571 F.3d 935, 946 (9th Cir. 2009) (“The overarching focus remains whether trial by class representation would further the goals of efficiency and judicial economy.”). By contrast, resolving these common issues once in the context of a class action will be much more efficient and fair. Superiority is met here.

c. **In the Alternative, The Court Should Certify the Alternative Classes or Certify an Issue Class Under Rule 23(c)(4).**

In the event that the Court declines to certify the Pirated Books Class, Plaintiffs submit that the Court should certify two proposed alternative classes based on datasets that Anthropic acquired and used in the same way ( [REDACTED] for Alternative Class #1, and [REDACTED] and [REDACTED] for Alternative Class #2<sup>6</sup>). Anthropic procured the datasets of books in each Alternative Class from the same pirated source, without a license or permission. Alternative Class #1 would consist of owners of the [REDACTED] copyrighted works in Anthropic’s [REDACTED] dataset, [REDACTED]. Miller Ex. 7. Alternative Class #2 would consist of owners of copyrighted works from Anthropic’s [REDACTED] and [REDACTED] datasets, [REDACTED]. Miller Ex. 4 at 9–10.

As a further alternative, a district court may, under Rule 23(c)(4), certify an issue class if “the adjudication of the certified issues would significantly advance the resolution of the underlying case.” *Valentino*, 97 F.3d at 1229. Issue class certification does not require predominance. *Id.* at 1234; *see also Cai v. CMB Exp. LLC*, 2024 WL 2334509, at \*12–13 (C.D. Cal. Apr. 4, 2024) (noting the “now widespread consensus” on this issue).

**VII. APPOINTMENT OF CLASS COUNSEL**

Plaintiffs request that Lieff Cabraser Heimann & Bernstein, LLP and Susman Godfrey L.L.P. be appointed as Class Counsel for this action. Fed. R. Civ. P. 23(g)(1) provides that “a

<sup>6</sup> Plaintiffs continue to discover additional pirated datasets as discovery unfolds, including as recently as March 20. Should discovery reveal additional pirated datasets, they may require further alternative classes.

1 court that certifies a class must appoint class counsel” and instructs the Court to consider:

2 (i) the work counsel has done in identifying or investigating potential claims in the  
3 action; (ii) counsel’s experience in handling class actions, other complex litigation,  
4 and the types of claims asserted in the action; (iii) counsel’s knowledge of the  
5 applicable law; and (iv) the resources that counsel will commit to representing the  
6 class.

7 Proposed Class Counsel possess the experience, ability, and resources to prosecute these  
8 claims and should be appointed as Class Counsel for this action. First, they have invested time  
9 and effort into preparing this case. Plaintiffs’ Counsel have conducted thorough pre-filing  
10 investigations, conducted extensive written discovery, overseen data inspections, presented a  
11 technology tutorial to the Court, argued motions relating to case sequencing and discovery,  
12 defended the Class Representatives in depositions, taken an early 30(b)(6), and have performed  
13 the important task of moving for class certification. Declaration of Rachel Geman and Justin A.  
14 Nelson in Support of Plaintiffs’ Motion for Appointment as Co-Lead Class Counsel, ¶¶ 7–22.

15 Second, Proposed Class Counsel have unmatched experience and expertise litigating class  
16 action cases generally—including (as relates to knowledge of applicable law) cases involving  
17 allegations that LLM companies have violated the Copyright Act<sup>7</sup> and harmed writers. *Id.* at  
18 ¶¶ 23–36, 44–55. For example, Proposed Class Counsel, with Cowan Debaets, serve as court-  
19 appointed interim Class Counsel in other pending copyright class actions against OpenAI and  
20 Microsoft for similar conduct as at issue here. *Id.* at ¶¶ 23–26.

21 Finally, Proposed Class Counsel have already invested significant resources in litigating  
22 this action on behalf of the Proposed Classes and are prepared to continue to do so through the  
23 resolution of this case. *Id.* at ¶¶ 7–22.

## 24 **VIII. CONCLUSION**

25 For the foregoing reasons, the Court should certify the Pirated Books Class and the  
26 [REDACTED] Books Class under Rules 23(b)(2) and 23(b)(3); appoint Plaintiffs as Class  
27 Representatives; and appoint Plaintiffs’ counsel as Class Counsel.

28 <sup>7</sup> Proposed Class Counsel are co-counseling in this case, and other similar matters, with copyright  
boutique Cowan Debaets Abrahams & Shepherd LLP.



1 Dated: March 27, 2025

Respectfully submitted,

2  
3 /s/ Rachel Geman

Rachel Geman

4 Rachel Geman (*Pro Hac Vice*)  
5 Jacob S. Miller (*Pro Hac Vice*)  
6 Danna Z. Elmasry (*Pro Hac Vice*)  
7 **LIEFF CABRASER HEIMANN**  
8 **& BERNSTEIN, LLP**  
9 250 Hudson Street, 8th Floor  
10 New York, New York 10013-1413  
11 Telephone: (212) 355-9500  
12 rgeman@lchb.com  
13 jmill@lchb.com  
14 delmasry@lchb.com

15 Daniel M. Hutchinson (SBN 239458)  
16 Reilly T. Stoler (SBN 310761)  
17 **LIEFF CABRASER HEIMANN**  
18 **& BERNSTEIN, LLP**  
19 275 Battery Street, 29th Floor  
20 San Francisco, CA 94111-3339  
21 Telephone: (415) 956-1000  
22 dhutchinson@lchb.com  
23 rstoler@lchb.com

24 *Proposed Co-Lead Counsel*

25 Justin A. Nelson (*Pro Hac Vice*)  
26 Alejandra C. Salinas (*Pro Hac Vice*)  
27 Collin Fredricks (*Pro Hac Vice*)  
28 **SUSMAN GODFREY L.L.P.**  
1000 Louisiana Street, Suite 5100  
Houston, TX 77002-5096  
Telephone: (713) 651-9366  
jnelson@susmangodfrey.com  
asalinas@susmangodfrey.com  
cfredricks@susmangodfrey.com

Rohit D. Nath (SBN 316062)  
**SUSMAN GODFREY L.L.P.**  
1900 Avenue of the Stars, Suite 1400  
Los Angeles, CA 90067-2906  
Telephone: (310) 789-3100  
RNath@susmangodfrey.com

Jordan W. Connors (*Pro Hac Vice*)  
**SUSMAN GODFREY L.L.P.**  
401 Union Street, Suite 3000  
Seattle, WA 98101  
Telephone: (206) 516-3880  
jconnors@susmangodfrey.com

J. Craig Smyser (*Pro Hac Vice*)  
**SUSMAN GODFREY L.L.P.**  
One Manhattan West, 51st Floor,  
New York, NY 10019  
Telephone: (212) 336-8330  
csmyser@susmangodfrey.com

*Proposed Co-Lead Counsel*

Scott J. Sholder (*Pro Hac Vice*)  
CeCe M. Cole (*Pro Hac Vice*)  
**COWAN DEBAETS ABRAHAMS  
& SHEPPARD LLP**  
60 Broad Street, 30th Floor  
New York, New York 10010  
Telephone: (212) 974-7474  
sshoulder@cdas.com  
ccole@cdas.com

*Additional Counsel for the Class*

**CERTIFICATE OF SERVICE**

I hereby certify that a true and correct copy of the foregoing and all documents attached hereto were served March 27, 2025 upon counsel of record via service by ECF.

/s/ Rachel Geman  
Rachel Geman

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28